From ARQMath 2020 to 2021

Topics in the scope

Dávid Lupták

Math Information Retrieval Research Group,
Faculty of Informatics, Masaryk University

`https://mir.fi.muni.cz/`

March 11, 2021

ARQMath Overview

Task 1: Answer Retrieval

- Given a posted question as a query, search all answer posts and return relevant answer posts.

Query

How can I evaluate $\sum_{n=0}^{\infty}(n+1)x^n$?

Asked 8 years, 5 months ago Active 4 months ago Viewed 34k times

384

How can I evaluate

$$\sum_{n=1}^{\infty}\frac{2n}{3^{n+1}}$$

146

I know the answer thanks to Wolfram Alpha, but I'm more concerned with how I can derive that answer. It cites tests to prove that it is convergent, but my class has never learned these before so I feel that there must be a simpler method.

In general, how can I evaluate

$$\sum_{n=0}^{\infty}(n+1)x^n?$$

sequences-and-series convergence power-series faq

edited Sep 24 '17 at 12:09 Parcly Taxel 51.7k ● 13 ■ 80 ▲ 120

asked Apr 3 '11 at 21:41 Backus 2,072 ● 3 ■ 12 ▲ 8

Search Results

1

No need to use Taxol series, this can be derived in a similar way to the formula for geometric series. Let's find a general formula for the following sum:

$$S_m = \sum_{n=1}^{m} nr^n.$$

...

2

It is equivalent to $x(x+1)(x+5)(x+6)+96=0$

Now

...

$$(x^2+6x)(x^2+6x+5)+96=0$$

3

If you want a solution that doesn't require derivatives or integrals, notice that

$$1+2x+3x^2+4x^3+\cdots = 1+x+x^2+x^3+\ldots$$
$$+x+x^2+x^3+\ldots$$
$$+x^2+x^3+\ldots$$

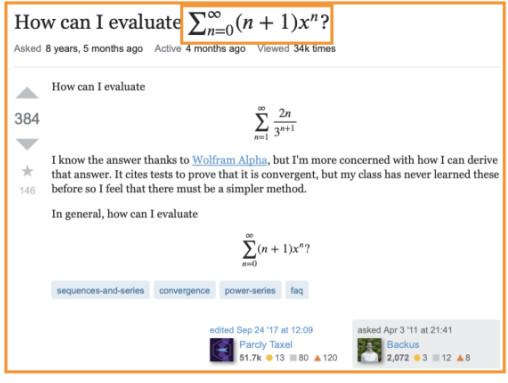
...

.
.
.

Task 2: Formula Retrieval

- Given a question post with an identified formula as a query, search all question and answer posts and return relevant formulas with their posts.

Query



Search Results

1 $\sum_{n=0}^{\infty}(n+1)x^n$

2 $\sum_{n=0}^{\infty}(n+1)x^n$

3 $\int_0^1 \frac{\ln(x+1)}{x^2+1}dx$

Topics (questions)

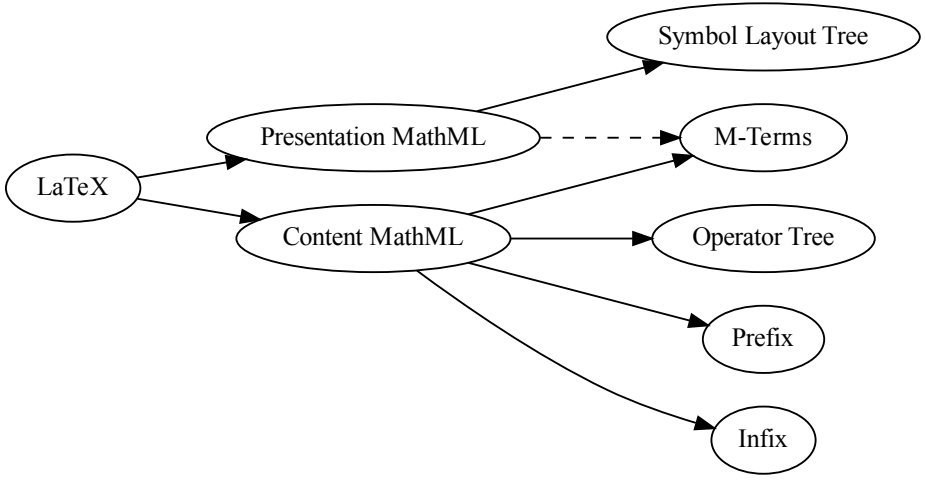
- 77 topics for Task 1
  - from various domains (real analysis, calculus, linear algebra, discrete mathematics, set theory, number theory, etc.)
  - categorized as computation (26), concept (10), proof (41)
  - the difficulty level spanned from easy problems (32), medium (21) to hard (24)
  - dependency on surrounding text (13), formulas (32) or both (32)
- 45 topics for Task 2
  - mathematical formulae selected from the topics from Task 1
  - criteria: complexity, elements, and text dependence

MIRMU Overview

Methods

Math Representations

- In our MIR systems, we used the following math representations:

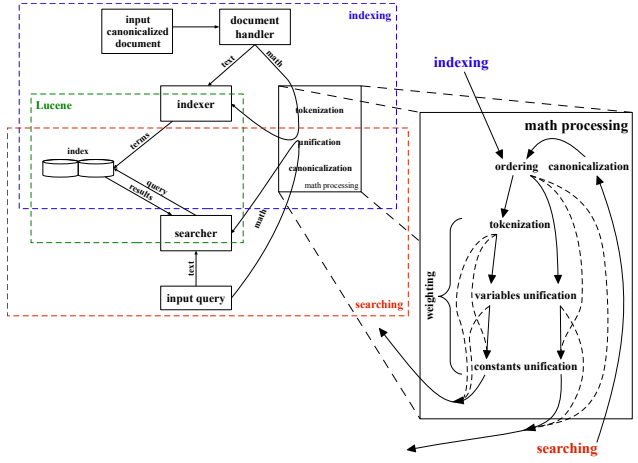


Methods

Corpora, Relevance Judgements, and Evaluation Measures

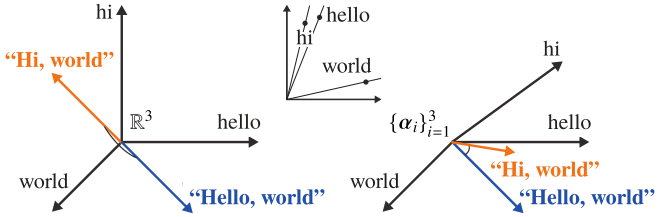
- For training, we used the following two corpora:
  1. ArXMLiv (four different subsets), [2] and
  2. Math StackExchange.
- For validation, we used the following two sets of relevance judgements:
  1. Automatic (param. opt., model sel.), and
  2. Human-Annotated (perf. est.).
- In our evaluation, we used the following two measures:
  1. Normalized Discounted Cumulative Gain Prime (nDCG'), [7] and
  2. Spearman's Correlation Coefficient (ρ).
- For retrieval, we used a machine with with 32 CPUs and 252 GiB RAM.
- For training embeddings, we used an NVIDIA GTX2080 Ti GPU with 11 GiB VRAM.

Math Indexer and Searcher (MIaS)



- Historically the first MIR system deployed in a digital mathematical library. [9]
- Uses TF-IDF with M-Terms extracted from CMML as a math representation.
- **Accuracy:** nDCG' 0.155, insignificantly below the Tangent-S baseline.
- **Speed:** avg. 1.24 s/topic, min. 0.1 s/topic, max. 7.27 s/topic.

Soft Cosine Measure (SCM)

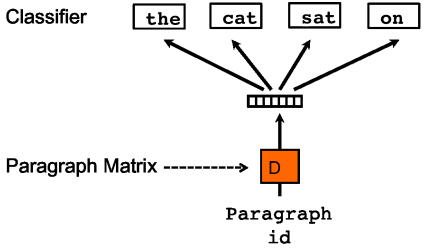


- Uses joint fastText [1] word embeddings of text & math to measure relatedness.
- Uses TF-IDF with the Prefix math representation and SCM [8, 4, 5] doc. similarity.
- Uses automatic relevance judgements to optimize parameters of fastText and SCM.
- Four different fastText models were trained:

1. Tiny (5 epochs, alternative submission)
2. Small (10 epochs, primary submission)
3. Medium (2 epochs on all corpora)
4. Large (10 epochs on all corpora)

- **Accuracy:** nDCG' 0.224 (small), insignificantly below the Approach0 baseline.
- **Speed:** avg. 58.46 s/topic, min. 30.52 s/topic, max. 502.84 s/topic.

Formula2Vec

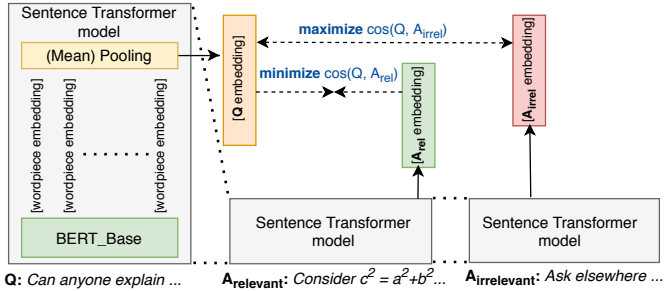


- Uses Doc2Vec DBOW [3] with the Prefix math representation and cosine doc. sim.
- Uses the optimal parameters of fastText and RedHat defaults for Doc2Vec.
- Four different Doc2Vec models were trained:

1. Tiny (5 epochs on no_problem ArXMLiv)
2. Small (10 epochs, alternative sub.)
3. Medium (2 epochs on all corpora)
4. Large (10 epochs on all corpora)

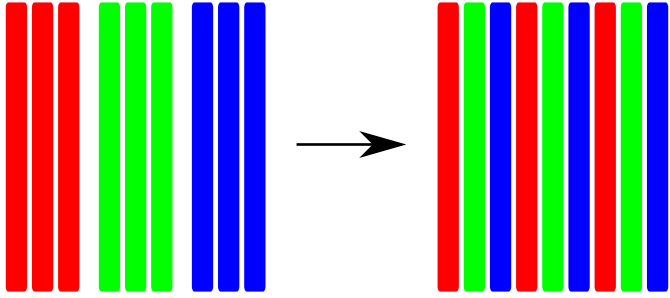
- **Accuracy:** nDCG' 0.050 (small), on par with DPRL and zbMath systems.
- **Speed:** avg. 3.23 s/topic, min. 3.14 s/topic, max. 7.87 s/topic.

CompuBERT



- Uses sBERT [6] with the LaTeX math representation and the cosine similarity.
- Uses our automatic relevance judgements to optimize the Triplet objective.
- Stark difference in performance between automatic and human-annotated r.j.'s.
- **Accuracy:** nDCG' 0.009, not significantly better than zero.
- **Speed:** avg. 3.43 s/topic, min. 3.2 s/topic, max. 3.67 s/topic.

Ensemble



- Interleaves the result lists of primary submissions: MIaS, SCM, and CompuBERT.
- Uses a parameter-free ensembling algorithm that only uses ranks, not scores.
- Results are ranked by median rank, then by frequency, and then interleaved.
- **Tie-breaking:** More than 40% of all results were arbitrarily interleaved.
- **Accuracy:** nDCG' 0.238, best of our systems, significantly better than all but SCM. The ensemble of all non-baseline primary submissions (0.419) best in competition.

Tangent-L Overview

Methodology

- Conversion – a “bag” of formulae and keywords
- Searching – Tangent-L to query the indexed corpus (MSE question-answer pairs)
- Re-ranking – Re-order the best matches by considering additional metadata
  - similarity
  - tags
  - votes
  - reputation

Results

- strong performance for topics that rely heavily on formulae
- strong at Computation-type and Proof -type topics, but is particularly weak at Concept-type topics
  - none of the Concept-type topics have a Formula-dependency
- excels at all three levels of difficulty: Easy, Medium, and Hard
  - topics relying on formulae (Formula-dependency or Both-dependency)

Bibliography

Bibliography I

[1] Piotr Bojanowski et al. “Enriching word vectors with subword information”. In: *Transactions of the Association for Computational Linguistics* 5 (2017), pp. 135–146.

[2] Deyan Ginev. *arXMLiv:08.2019 dataset, an HTML5 conversion of arXiv.org*. SIGMathLing – Special Interest Group on Math Linguistics. 2019. URL: https://sigmathling.kwarc.info/resources/arxmliv-dataset-082019/.

[3] Quoc V. Le and Tomas Mikolov. “Distributed Representations of Sentences and Documents”. In: *CoRR* abs/1405.4053 (2014). URL: http://arxiv.org/abs/1405.4053.

[4] Vít Novotný. “Implementation Notes for the Soft Cosine Measure”. eng. In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18)*. Torino, Italy: Association for Computing Machinery, 2018, pp. 1639–1642. ISBN: 978-1-4503-6014-2. DOI: 10.1145/3269206.3269317.

Bibliography II

[5] Vít Novotný et al. *Text classification with word embedding regularization and soft similarity measure*. 2020. arXiv: 2003.05019 [cs.IR]. URL: https://arxiv.org/abs/2003.05019.

[6] Nils Reimers and Iryna Gurevych. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks". In: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 3982–3992. DOI: 10.18653/v1/D19-1410. URL: https://www.aclweb.org/anthology/D19-1410.

[7] Tetsuya Sakai and Noriko Kando. "On information retrieval metrics designed for evaluation with incomplete relevance assessments". In: *Information Retrieval* 11.5 (2008), pp. 447–470.

Bibliography III

[8] Grigori Sidorov et al. “Soft similarity and soft cosine measure: Similarity of features in vector space model”. In: *Computación y Sistemas* 18.3 (2014), pp. 491–504.

[9] Krzysztof Wojciechowski et al. *The EuDML Search and Browsing Service – Final*. Deliverable D5.3 of EU CIP-ICT-PSP project 250503 EuDML: The European Digital Mathematics Library, revision 1.2 `https://project.eudml.eu/sites/default/files/D5_3_v1.2.pdf`. Feb. 2013.

MUNI
FACULTY
OF INFORMATICS